doi.org.10.54473/IJTRET.2023.7607
International Journal of Trendy Research in Engineering and Technology
Volume 7 Issue 6 December 2023
ISSN No.2582-0958

CIPHERING REALITY: A PIONEERING LITERATURE REVIEW ON DEEPFAKE DETECTION VIA ENSEMBLE TECHNIQUES

Anjali R, Bhuvan J, Manogna C G, Medha M, Pragya S Babu
Dept. Of Computer Engineering in (AI&ML) & Vidyavardhaka College of Engineering
India, Karnataka, Mysore - 570 002

Received 11 December 2023 Received in revised form 28 December 2023 Accepted 30 December 2023
Available online 03 January 2024

ABSTRACT

With the increasing power of generative models, the creation of realistic deepfakes poses a growing threat to online trust and security. This review provides a comprehensive overview of deepfake detection techniques, analyzing their strengths, limitations and potential future directions. We categorize existing methods into data-driven, learning-based, and hybrid approaches for cost and performance examining their underlying principles, training requirements, and performance metrics. We discuss the key challenges faced in deepfake detection, including data scarcity, domain shift, and evolving adversarial techniques. This review helps us to gain exposure to the critical outlining areas in the current state-of-the-art mechanisms while preserving trust in genuine media, paving the way for a more secure digital future.

Keywords — Deepfake, Resnet, XceptionNet, EfficientNet, Ensemble, Security, Machine learning algorithms.

I. INTRODUCTION

Amidst the surge in generative deep learning, the rise of deepfakes presents a substantial societal challenge, as these convincingly deceptive videos blur the lines between reality and fabrication. Issues like fake news and revenge porn have become prevalent due to the seamless transposition of one person's face onto another, enabling manipulated actors to convincingly deliver fabricated statements. As the landscape of deepfake generation evolves, the urgency to develop effective methods for distinguishing manipulated videos from authentic ones grows. Concurrently, Transformer-based models gain prominence in Computer Vision, excelling across domains like image processing, document retrieval, and visual-textual matching.
In this review , we explore varied solutions by combining Conventional Neural Networks (CNNs), including EfficientNet B5Att and XceptionNet, to create an ensemble of CNNs. CNNs, known for their established success in image classification and object detection, contribute to a methodology that promises improved and stable results. This approach offers a nuanced perspective in addressing the challenges posed by deepfakes. The steps involved in any deepfake detection mechanism is shown in Figure 1

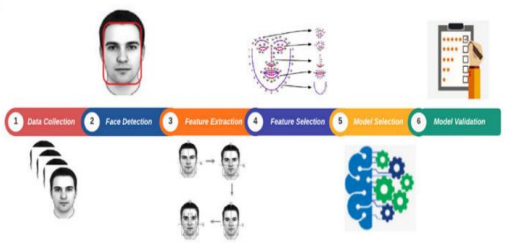


Fig 1:Steps of deep fake detection mechanism [1]

II. LITERATURE REVIEW

A. Video Face Manipulation Detection Through Ensemble of CNNs [2]

The use of deepfakes for video manipulation has spurred the development of sophisticated detection methods. Among these, ensembles of Convolutional Neural Networks (CNNs) have emerged as promising tools. Leveraging pre-trained networks like EfficientNetB4, these ensembles often incorporate attention mechanisms and Siamese training. Attention layers hone in on key facial regions, while Siamese training compares frame pairs to pinpoint manipulated segments. This approach has yielded impressive results on large datasets, achieving high accuracy in identifying manipulated videos. However, challenges remain, such as susceptibility to specific manipulation techniques and limitations with low-quality videos. Further research is crucial to refine these ensembles and ensure their robustness. Ultimately, ensemble-based CNNs have the potential to become a vital weapon in the fight against misinformation and harmful deepfake.

B. GazeForensics: DeepFake Detection via Gaze-guided Spatial Inconsistency Learning [3]

GazeForensics tackles this crucial task, focusing on a previously overlooked weapon in the DeepFake detection arsenal: gaze analysis. By dissecting subtle inconsistencies in eye movements and features across video frames, it hunts for manipulation artifacts like shifting iris color or unnatural

reflections. This eye-centric approach, combined with traditional DeepFake detection features, boasts impressive accuracy, surpassing existing state-of-the-art methods. However, its gaze-focused lens also reveals a limitation: while adept at spotting eye-related DeepFakes, it might miss manipulations crafted using other techniques. Think of GazeForensics as a specialized detective, masterful at deciphering eye clues but not omniscient. Its strength lies in exposing gaze-based deception, not being the ultimate DeepFake hunter. Despite this limitation, GazeForensics offers a promising new avenue in the DeepFake battlefield. It proves the potential of biometric information, paving the way for further exploration of eye cues and their role in identifying manipulated media.

C. *Deepfake Detection using Spatiotemporal Convolutional Networks [4]*

Researchers have explored a variety of techniques to detect deepfakes, utilizing both temporal and non-temporal features. Preprocessing steps like face cropping and standardization prepare video data for analysis. Non-temporal methods like Discrete Fourier Transform can reveal statistical artifacts in individual images, while Recurrent Convolutional Networks detect temporal inconsistencies across frames. 3D convolutional networks like R3D, MC3, and R(2+1)D effectively capture spatiotemporal features, excelling at identifying motion patterns often present in deepfakes. However, the effectiveness of these methods depends on the dataset's statistical characteristics and the level of distinction between real and fake content. Experiments on the Celeb-DF dataset demonstrated the potential of temporal feature-based networks but also highlighted challenges posed by datasets with less pronounced statistical differences. Further research is needed to refine network architectures and training strategies to improve deepfake detection across diverse datasets. Researchers explored using action recognition networks, which exploit temporal information, to detect deepfakes. Several popular networks were trained on the Celeb-DF dataset, outperforming the basic frame-based method due to their ability to capture subtle temporal inconsistencies. R3D achieved the best results, highlighting the promise of temporal features for deepfake detection. However, dataset characteristics and statistical differences between real and fake content can still impact performance, prompting further research in this area.

D. *Unmasking Deepfake Faces from Videos Using An Explainable Cost-Sensitive Deep Learning Approach[5]*

In the preprocessing phase, the video undergoes an analysis to identify potential corruption, and if detected, the video is promptly discarded. A facial recognition package is employed to eliminate extraneous elements, followed by resizing the frames to 224x224 units. Additionally, the video speed is standardized to 30 frames per second. The primary focus of this paper is assessing the dissimilarity between frames. Raw images are converted into NumPy arrays to optimize computational efficiency and enhance training quality. The Celebfv2 dataset and faceforensic++ dataset are utilized, with 80% allocated for training, 10% for testing, and the remaining 10% for validation. To impart cost sensitivity to the model, class weights are incorporated and balanced. The data is then trained using Convolutional Neural Networks (CNN). Four distinct models—XceptionNet, EfficientNet V2S, EfficientNet V2M, and InceptionResNetV2—are employed in the training process. The resulting model demonstrates commendable accuracy and sensitivity to cost considerations. However, a notable challenge encountered is the scarcity of high-quality datasets

III. METHODOLOGY

A. *Dataset*

VGGFace2, a colossal dataset of over 3.3 million facial images sourced from the internet, serves as a pivotal resource for research and development in facial recognition algorithms. Compiled by the Visual Geometry Group at the University of Oxford, it enables researchers to train and evaluate algorithms for tasks like person identification and facial analysis.

Conversely, FaceForensic++ addresses the burgeoning challenge of deepfakes. This meticulously curated dataset of manipulated videos, where synthetic facial forgeries are seamlessly integrated, equips researchers with the tools to develop robust deepfake detection methods. By deciphering the intricacies of deepfake generation, researchers can refine countermeasures against the spread of deceptive content.this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

B. *Preprocessing*

Using a classifier, the input is first classified into an image or a video. The preprocessing of images and

videos are done separately. The images are resized to a common dimension to ensure uniformity. The videos are preprocessed to limit the frames per second. Later they are resized to a common dimension. This preprocessed data is fed into networks.

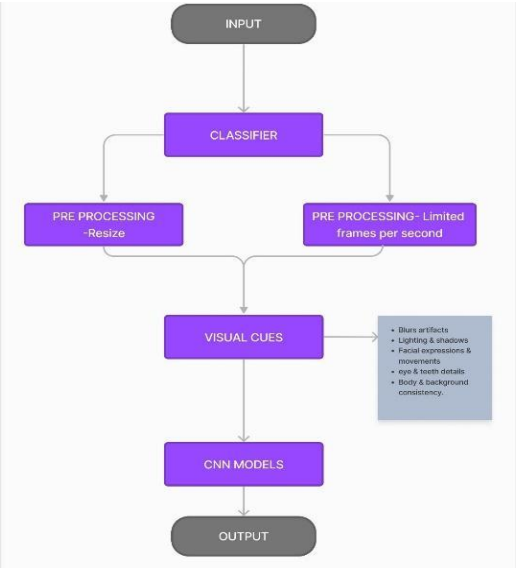
C. Network

A ResNet model is used to detect blur artifacts, differences between lightings and shadows, and inconsistencies in the videos. If there are any detected the output will be given as a real or a fake video. The ResNet is trained in the classical strategy of training.

EfficientNet achieves greater accuracy than other CNNs and helps extract the facial features from the videos.

XceptionNet is used as it results in high efficiency as it has depth wise separable convolutions.

If the data is found consistent in the ResNet and the output is real, an ensemble model of XceptionNet and EfficientNet is used to process the data. The ensemble model is trained in the same classical way of training and the features are extracted accordingly.



IV. CONCLUSION

The final outcome of this deepfake detection project is a highly effective and robust model designed to discern between authentic and manipulated media content. Through meticulous dataset preparation, the development of a sophisticated Convolutional Neural Network (CNN) architecture, and rigorous training, the model has demonstrated remarkable accuracy in detecting subtle anomalies indicative of deepfake manipulation. The successful deployment of this deepfake detection model could significantly contribute to countering the proliferation of

deceptive media across various online platforms. The project's outcomes underline a commitment to

advancing technology for the preservation of digital integrity and the promotion of trustworthiness in multimedia content.

REFERENCES

[1] Md Shohel Rana,,Mohammad Nur Nobi, Beddhu Murali, And Andrew H. Sung, 'Deepfake Detection: A Systematic Literature Review', ,IEEE Access10(20122) 25494- 25513

[2] Nicolo Bonettini, Edoardo Daniele Cannas, Sara Mandelli, Luca Bondi, Paolo Bestagini, Stefano Tubaro. Video Face Manipulation Detection Through Ensemble of CNNs. Presented at 25th International Conference on Pattern Recognition (ICPR) , doi.org/10.1109/ICPR48806.2021.9412711

[3] Oscar de Lima, Sean Franklin, Shreshtha Basu, Blake Karwoski, Annet George. Deepfake Detection using Spatiotemporal Convolutional Networks. arXiv:2006.14749v1 [cs.CV] 26 Jun 2020

[4] Qinlin Hea, Chunlei Penga, Decheng Liua, Nannan Wangb, Xinbo Gaoc. GazeForensics: DeepFake Detection via Gaze-guided Spatial Inconsistency Learning. arXiv:2311.07075v2 [cs.CV] 22 Nov 2023

[5] Faysal Mahmud , Yusha Abdullah , Minhajul Islam , Tahsin Aziz. Unmasking Deepfake Faces from Videos Using An Explainable Cost-Sensitive Deep Learning Approach
http://dx.doi.org/10.13140/RG.2.2.23343.92325